assigned to the target is converted into a standardized average rating score for the target (SAR score).

The distribution of the sum of ratings for the controls can be considered as the distribution of ratings associated with that condition. Reduced to the level of individual trials we assume this distribution to be typical for the condition and express all ratings in this distribution of average ratings. Thus, all ratings are converted into standard normal scores by computing its distance from the mean of average ratings for the controls of the trials and dividing it by the standard deviation observed for these average ratings.

Then for each trial a SAR score for the target is defined as the difference between this standard normal score for the target and the average standard normal score for target and controls. Since the SAR scores are based on true standard normal scores, which means scores obtained from a normal distribution, SAR scores can be considered normal too. For each trial the sum of SAR scores for controls and targets is zero. Therefore, in the case of related samples we might compare individual achievement over conditions by calculating a product-moment correlation between the SAR scores of the two conditions.

Although the randomization test described above seems statistically sound we further studied its properties, especially regarding its sensitivity to detect ESP. To this end we conducted a computer simulation of 100 "experiments" for each combination of two variables. Each experiment consisted of 20 trials and 5 pictures per trial and was simulated by randomly generating 20 rows of 5 numbers between rating values 0 and 30, inclusive. The two variables involved were subjects' rating behavior and amount of ESP. For rating behavior we manipulated the probability of selecting rating values of zero. The amount of ESP was operationalized as the number of subjects assigning the highest rating value to the target in addition to what could be expected by chance.

From the data obtained it can be concluded that in most conditions the sensitivity of the SSR scores is rather low and less than that when, for instance, a simple binomial test was applied. Only in extreme cases of rating behavior and amount of ESP do the SSR scores become more sensitive than the binomial test. For instance, in the case of 5 ESP hits when in total 5 + 15/5 = 8 hits can be expected, the binomial yields an exact one-tailed probability of p = .01 whereas the SSR score yields on average a Z of 1.7 with an associated one-tailed probability of .045.

In the same simulation studies Stanford Z-scores were computed. We know that the distributions for these Z-scores are non-normal but leaving this aside we found that in most cases the sensitivity of t-test evaluations based on Stanford Z-scores is comparable to that of evaluations based on SSR scores. However, SSR

scores appear more sensitive than Stanford Z-scores in cases of strong ESP and extreme rating behavior.

From these findings some practical conclusions can be drawn. In general we must assume that the ESP influence on the data is relatively little. Hence, unless there is reason to expect a strong ESP influence in the experiment the binomial test can be assumed to be more sensitive than an evaluation based on the rating values. The same applies for experiments in which no extreme rating behavior can be expected, for instance, in an experiment in which an atomistic approach to the judging is followed. In that case we expect in general nonzero ratings assigned to all pictures, and our findings show that in that case the SSR scores, as well as Stanford's Z-scores, are rather insensitive.

A METHODOLOGY FOR THE DEVELOPMENT OF A KNOWLEDGE-BASED JUDGING SYSTEM FOR FREE-RESPONSE MATERIALS

Dick J. Bierman (Dept. of Psychology, University of Amsterdam)

It has been found that certain judges perform consistently better than others when matching targets to a target set. It seems unlikely that this is purely because of the judge's psi, since psi generally does not display consistent behavior. Therefore, it might be hypothesized that it is the (intuitive) knowledge of the specific judge that accounts for his better performance on this task. It has been proposed (Morris, EJP, 1986, 137-149) that the use of expert systems might help psi researchers in tasks where they lack expertise, such as in the detection of fraud. Morris argues that the expertise of magicians could be formalized in such a system and made available to each individual researcher. Similarly, the expertise of the best judges of free-response material could become available through implementation of a knowledge-based free-response judging system. This use of techniques from the field of artificial intelligence (AI) to represent scarce knowledge should not be confused with the use of AI techniques for the representation of free-response material (Maren, RIP 1986, 97-99). According to Maren, the free-response material and the protocols should be represented in the form of trees in which the nodes are perceivable "objects," like "flames," and the links represent relations, like "adjacent to." We expect that focusing our attention on the (knowledge used in the) human matching process might reveal more fundamental information about the role of the meaning of the material. It is striking that in Maren's proposed representation of complex target material only visual features are present. Actually, the type of visual matching that Maren proposes to be done by a machine can be better performed by any sighted human.

It should be remarked that the crucial element in the development of expert systems nowadays is not the implementation of the system but the elicitation of the knowledge that has to be entered into the system. In the case of knowledge about trickery, for instance, it is doubtful that one can find experts who are willing to transfer their knowledge. Apart from that, the detection of trickery is largely driven by visual information. The proper representation of this visual knowledge might also be a major problem in this domain of expertise. In the case of free-response judging one can expect cooperation from the expert judges. Although the material is also visual there are strong indications that simple key words are able to represent these pictures satisfactorily. This conclusion can be drawn from the analytical judging procedures developed by Jahn et al. (Jahn et al., JP, 1980, 207-231).

Analytical judging versus knowledge-based judging. It has been found that simple (linear) regression formulas make predictions comparable to or better than human experts in the domain of psychodiagnostics. Thus, it is not surprising that the analytical judging procedure very similar to an approach by linear regression also yields satisfactory results. However, it should be noted that although its average performance is adequate, this approach fails in pathological cases. It appears that this is because of the failure to take into account any interaction between the predictor variables. In the analytical judging procedure, for instance, the simultaneous occurrence of two elements is counted as the sum of the scores for the cases when they occur alone. Thus, if two elements together have a symbolic meaning that is not contained in either element separately, this meaning is missed in the analytical judging procedure. A knowledge-based judging system is capable of representing and using this type of knowledge.

Matching as classification task. Most problem-solving tasks can be seen as classification tasks. In the case of matching free-response material from psi experiments, however, there is a special problem. Since the categories "correct match" and "incorrect match" in psi research are determined by chance, these categories do not have objective attributes. Thus, the task cannot be modeled as a direct classification task. Therefore, we propose to model the matching process as a double classification process. The judge is thought to begin with a classification of the protocol in one of his internalized categories. Secondly, this procedure is repeated for each of the members of the target set. Finally, the results of these classifications are evaluated using overlap measures. If no clear-cut match can be made a secondary evaluation is done which takes into account (subtle) interactions among attributes.

Knowledge-elicitation methods. The elicitation of knowledge needed to drive expert systems is a "bottleneck problem." This was one of the reasons to simulate the research in machine-learning methods as a means of explicating knowledge. Very often the rather

unstructured interview approach is accompanied by so-called rapid prototyping. This means that the system is implemented while the knowledge base is essentially of low quality and incomplete. This might result in poor final systems, like most rule-based systems to date. If this is already the case for rather well-understood areas of human expertise it seems unwise to use an unstructured elicitation procedure for the expertise of free-response judging. In more structured approaches emphasis is given to the necessity of a well-specified framework for interpretation of the verbal material, be it interviews with, or thinking aloud protocols produced by, the expert. In the present paper it is proposed to combine the structured knowledge-elicitation procedure with the use of learning systems.

Proposed procedure. The proposed methodology differs from accepted methodologies by using information already present in the data base of classified cases. The elicitation procedure consists of three major parts: (1) Learn, (2) Pathology detection, (3) Confrontation.

In the first phase the expert judge will be interviewed on the set of attributes that are used to describe a target picture. Also, the primary set of classes is formulated. After that, a training set of old cases is selected to be presented to a learning system. Each case consists of a series of attribute values together with the classification by the expert judge. After the training the systems are able to classify other cases from the old data base and to compare classifications of the target set with the classification of the protocol. The trained system has become a (first-order) model of the expert judge.

In the second phase the remainder of the old data base is presented to the "trained" system for judging. If the judging by the system differs from that made in the past by the human expert, we call this a "pathological case."

In the third phase the human expert is confronted with the set of pathologies. The knowledge engineer might directly ask the expert why he or she deviated from the model or give him or her the cases to solve again while thinking aloud. Analysis of the thinking-aloud protocol should occur in terms of deviations from the model and thus produce additions to the knowledge base.

The automated concept learner. Previous work that tried to apply learning systems to the process of knowledge acquisition used systems like Automated Concept Learning System (ACLS), which construct a decision tree from examples. However it was found that although the resulting decision trees were able to classify new cases properly, these trees, which represent the knowledge of the human expert, very often were hardly recognized by the same expert. This decision-tree representation offered therefore not a fruitful framework for the knowledge engineer to base his or her further

Approved For Release 2000/08/11 : CIA-RDP96-00792R000400180001-4

interviews. This situation is not very different from a representation by linear regression models which have shown to have considerable predictive power. However, the linear regression formula does not make a lot of sense to the human expert. Therefore, we have proposed elsewhere not only to use an ACLS type of learning system but also to use a learning system that is supposed to create a psychologically valid representation of the human expert's knowledge.

The prototype learner. The "prototype" model has been developed by Rosch. In contrast with linear regression models, the "prototype" model allows for nonmonotonic relations between the values of the attributes and the class determination. So, apart from an implementation of a decision-tree building system à la ACLS, a system has been implemented that is capable of learning categories as proposed in the Rosch model. During the learn phase a training set of old cases, consisting of the values of the attributes and the resulting classification, are offered to the system. The system learns which attributes contribute to which degree to the final classification decision. After the learning phase new cases can be offered to the system which will calculate an overlap score of the new instance with the "prototype" of a class.

Concluding remarks. Current work by the present author using a similar knowledge-elicitation approach in the domain of psychodiagnostics is promising. It appears that "intuitive" knowledge can be elicited with the proposed approach and implemented as a moderator of a primarily pattern-matching-based classification.

NEW INTERPRETATIONS OF ESP LITERATURE*

A CRITICAL REVIEW OF THE DISPLACEMENT EFFECT

Julie Milton (Dept. of Psychology, University of Edinburgh, 7 George Square, Edinburgh EH8 9JZ, Scotland)**

The "displacement effect" in ESP research refers to a situation in which the percipient, instead of describing the intended target for a particular trial, describes some other experimental material. Despite the fact that over 100 papers have dealt with some aspect of the displacement effect since the effect caught the general interest of parapsychologists in 1940, no exhaustive review of the displacement literature has appeared. It was felt that such a review would be timely for a number of reasons, partly because the attitude of researchers these days to the apparent occurrence of displacement is generally one of irritation, whereas earlier researchers reacted with a more positive (and hence possibly more productive) interest; partly because recently some researchers have suggested that in the context of finding limits for psi, the circumstances under which displacement occurs and the extent to which displacement is a "deliberate" error or a genuine error on the part of the percipient may have some theoretical importance. Another reason for a review would be to examine the characteristics of displacement as a phenomenon of interest in itself.

In the past, researchers have explored two main lines of research with respect to displacement; the first has involved the possibility of a relationship between scoring on targets of different displacements, and the second, the possibility of a relationship between displaced scoring and psychological and situational variables.

Concerning the possibility of a relationship between scoring on targets of different displacements, there are a couple of potentially important statistical artifacts that apply to forced-choice studies which can give rise to the appearance of displacement

*Chaired by Erlendur Haraldsson.

**I am grateful to the Perrott-Warrick Studentship in Psychical Research for financial support during the writing of this paper.

Approved For Release 2000/08/11 : CIA-RDP96-00792R000400180001-4